Readme

**Ques.1 Satellite UCI dataset**

Uses standard Satellite UCI dataset. The code includes:

* Loading the dataset and perform splitting into training and validation sets with 70:30

ratio.

* Use tsne plot to visualise the dataset.
* Implement the kNN algorithm from scratch.
* find k using grid search.
* Plot the error vs number of neighbors graph (k) and Report the optimal number of neighbours.
* Report the training and the validation accuracy and classification report

**Ques.2 IRIS Dataset**

Uses the IRIS (UCI Machine Learning Repository: Iris Data Set) dataset. The code includes:

* Loading the dataset and perform splitting into training and validation sets with 70:30
* Implement the Kmeans algorithm using sklearn.
* Find the optimal number of clusters using the elbow method.
* Plot the error vs number of clusters graph and Report the optimal number of cluster found.
* Use Scatter plot to visualize the dataset to depict the clusters formed (optimal).
* Report the training and the validation accuracy and classification report.

**Usage:**

The given file contains 2 folders for Question 1 and Question 2 respectively. After unzipping the files, each folder contains a separate folder for the database used and the code file in .ipynb format.

The ipynb file can be viewed in Google Colab or in Jupyter Notebook for both the questions which includes software code, computational output, explanatory text and multimedia resources in a single document.

Finally a report file which contains all the necessary details about the process involved and analysis of the results.

**Note: All the code is written on Google Colab IDE, hence if this notebook is run on other platform like jupyter notebook or vs code, The path of the data file should also be changed according to the location of file as it is stored in my google drive originally, else it will show an error.**

**Members:**

**Yash Gandhi 2K18/CO/402**

**Ashi Gupta 2K18/SE/040**